Canada. Office of the Auditor General.
The evaluation function in research and development.

(Discussion paper no. 12)

Office of the
Auditor General
of Canada

Bureau du
vérificateur général
du Canada

DISCUSSION PAPER NO. 12

THE EVALUATION FUNCTION IN RESEARCH AND DEVELOPMENT

BY

J.R. Rutherford

FEBRUARY 1983

DISCUSSION PAPER SERIES

DOCUMENTS DE DISCUSSION

attached paper has been prepared to stimulate ght and discussion regarding our audit vities. The views expressed are those of the or and therefore should not be construed as e of the Office.

r comments would be appreciated and should be cted to the attention of the author.

tional copies of this paper, or other papers in "Discussion Paper Series", may be obtained gh the PROFESSIONAL PRACTICES GROUP.

Le document ci-joint vise à stimuler la réflexion et la discussion sur nos activités de vérification. Les opinions exprimées dans ce texte sont celles de l'auteur et, par conséquent, ne lient pas le Bureau.

Vos commentaires seraient appréciés et vous êtes priés de les faire parvenir à l'auteur.

Vous pouvez vous procurer des exemplaires supplémentaires de ce document ou des autres écrits de la série des "Documents de discussion" en vous adressant à la DIRECTION DES MÉTHODES PROFESSIONNELLES.

DISCUSSION PAPER NO. 12

THE EVALUATION FUNCTION IN RESEARCH AND DEVELOPMENT

BY

J.R. Rutherford

FEBRUARY 1983

The Evaluation Function in R&D

Introduction

In the 1981 annual Report of the Auditor General we reported on our government-wide audit of the management of R&D. At the time of the audit, a number of issues remained unresolved in the area of evaluation. Consequently, we stated the audit criterion, or management standard, for evaluation in deliberately vague terms. In this paper we describe a conceptual framework for evaluation in R&D on which future audits of evaluation in R&D could be based. In general, we consider that program evaluation should build on current practices which emphasize on-going, continuous evaluation by R&D program managers. However, there should also be periodic independent evaluations of R&D activities.

Periodic evaluations should focus, but not exclusively, on three aspects of the R&D process:

- the appropriateness and efficacy of the R&D management process;
- the appropriateness and quality of management and research personnel as well as facilities and equipment; and
- the quality of the research work being done.

Wherever possible and appropriate this focus should be expanded to include direct measurement of the impact and effects of R&D, but the major justification for focusing on the process rather than on results lies in the long time lags and uncertainty inherent in the R&D process. Recent R&D results, whether operational (e.g., papers published) or program (e.g., economic growth), are very often the outcome of research work, management decisions and conditions of 5 to 10 years earlier. Assessment of such results provides no "necessary,

logical" proof that current research activities will produce goal-related results in the next 5 to 10 years.

Principles For Developing Guidelines For Evaluation in R&D

In developing these guidelines, we have kept in mind two principles:

- the evaluation function should be tailored to both the management practices and the process being managed; and
- any new evaluation function should build on existing R&D management and evaluation practices.

The next sections describe the basis for our approach to satisfying the first principle. The need for the second principle became apparent during the government-wide audit, when we found there is much evaluative activity going on.

It is not expected that rapid acceptance will follow the adoption of any proposal on this subject. The literature on managing R&D is quite extensive but there is general agreement that "there is not a generally accepted body of knowledge that definitely establishes a methodology for objectively evaluating the subject, for quantitatively measuring the benefits derived from R&D, for relating results achieved to management practices, or for measuring the economic benefits from specific R&D activities".*

A major reason for this is that the word "evaluation" has several meanings in the research community. It is used for the project selection process as well as for the review of progress and results of project activities, in the sense of "pre-evaluation" and "post-evaluation", between which there is often not a

---

*Blake, S.P., Managing for Responsive R&D, W.H. Freeman & Company (1978), p. iii.

great deal of difference. Consider a program made up of several parallel research projects, or a research project made up of several research tasks. The pre-evaluation process involves an assessment of whether the people and facilities together are capable of producing a desired result, whether this result will make a contribution to achievement of program or project goals, and whether it will be possible to complete the project or tasks with the resources allocated. A post-evaluation is conducted periodically, or after the scheduled completion of the tasks or projects. Such evaluations usually also serve as pre-evaluations for program or project continuance.

There is a further complication in the federal government, where the terms "program evaluation" and "internal audit" have specific functional and organizational meanings. The R&D evaluation practices suggested in this paper require some of the skills of a program evaluator, a management auditor and a technical expert. We leave open the question of how this blend of skills should be administered - by the internal audit group, the program evaluation group or R&D management.

A Simple Model of R&D

Our approach to evaluation is based on a simple model of R&D and its management. The model highlights a key R&D activity: **linking a perception and a potential.** In greater detail, the two things to be linked are:

- the **perception of a need,** which R&D can make a contribution toward satisfying, or the **perception of an opportunity** for exploiting a technological or scientific capability; and

- the **potential capability** of satisfying the need or exploiting the opportunity. This capability has two aspects:

  (a) the scientific competence of the person or group to do the work; and

(b) the technical feasibility of satisfying the perceived need or opportunity, using the suggested technical approach.

Briefly, the model presents the view that the heart of R&D management involves linking a perceived need or opportunity to a capacity to solve the related problems. This linking occurs at the idea-generation and project selection stages as well as during strategic planning. It must, of course, be followed by the solving of the technical problems.

It is commonly said that achieving results from research and development activity can be uncertain, and this model indicates why. Some uncertainty lies in the technical assessment of whether a particular phenomenon or piece of technology can lead to the satisfaction of a need or opportunity. There is also uncertainty about whether available persons or groups have the capacity to solve the problem, given that it is hypothetically feasible to do so. Other sources of uncertainty are whether the perceived need or opportunity corresponds to a real need or opportunity and how much time and cost it will take to make the research idea ready for application.

Characteristics of R&D That Are Relevant to its Evaluation

R&D has certain characteristics that should influence the approach to its evaluation. There are usually long time lags between obtaining research results and seeing their effects. Uncertainty about producing a result and about its impact are often considerable. R&D is not hierarchical, so there is often no system of goals and objectives tightly linking tasks and projects. It is not always possible to aggregate the progress of individual projects to produce an appropriate indicator of program progress. The contribution of R&D results to supported programs is difficult to identify in many cases; research programs are typically heterogeneous; and R&D is usually best viewed as a means to an end, not as an end in itself.

Each of these points is elaborated in Appendix A.

Periodic Evaluation in R&D: Suggested Approach

The focus of periodic evaluations of R&D activity should be on the three aspects listed earlier, which are discussed in more detail here.

In regard to evaluation of the management process, a management review or audit of the strategic and operational planning system and the procedures for project selection, operational control and continuation review are commonly accepted as important. Such reviews serve to answer the question of whether R&D activities are being appropriately focused and efficiently conducted. The personnel staffing practices and the performance appraisal procedure can also be reviewed to determine their appropriateness.

The second aspect involves assessing the quality of management decisions as shown in strategic plans, projects selected, hiring and assignment of research personnel and acquisition of equipment. This usually involves outside experts who are periodically invited to review ongoing projects and activities to ascertain that the facilities and projects are as good as might be expected; the experts can also give an informed opinion on whether the perceived opportunities really exist. These evaluations are designed to find out if current management practices have been successful in matching people, projects and opportunities. Such reviews can also assess whether currently stated objectives continue to be appropriate and if R&D activities are likely to produce the desired results. The client for the research outputs will be involved in the latter aspects of such reviews.

The third aspect involves assessing the quality of the research activities. Outside experts can assess the validity of the approaches adopted for pursuing project objectives and the skill with which the work is being done. Reviewers can determine the quality of recent outputs using such measures such as professional judgement of publications, citation frequency and patents issued and licensed. Outputs of earlier years may be analysed for their socio-economic impact, but this is difficult because other activities make significant contributions, and attributing socio-economic impacts solely to the research contribution may not be justified.

In terms of the federal government's division between program evaluation and internal audit, the first issue could be handled by an appropriately experienced internal audit group, the second could be co-ordinated by either and the third by the program evaluation group.

Link to existing evaluation practices. Good R&D management makes evaluative judgements continuously - at annual reviews of strategy and at annual or semi-annual project selection or continuation reviews. One purpose of the periodic reviews is to assure senior management that these evaluative functions are being performed and performed well. Another could be to provide senior management with an independent view of organizational capacity and potential for R&D. However, some of the judgements and opinions developed by the outside experts will be highly technical and specific to particular research projects. Such information will be more useful to lower levels of research management than to senior management. Hence the client for periodic evaluation includes relatively junior levels of management.

Program evaluation. Consider now the four general classes of evaluation questions listed in the OCG Guide on Program Evaluation (p. 7):

- program rationale (does the program make sense?)
- impacts and effects (what has happened as a result of the program?)
- objectives achievement (has the program achieved what was expected?)
- alternatives (are there better ways of achieving the results?).

Program rationale and alternatives can often be investigated simultaneously. Our R&D management model is at the heart of the question of rationale: is R&D activity likely to make an adequately valuable contribution? Another question to be addressed by the evaluator is whether there is a need or opportunity

that requires R&D for its satisfaction. Answering this would involve outside experts as well as the client for R&D results. Technological forecasting of sectoral needs or opportunities as well as consideration of the sector maturity (its history of technological innovation) are often required. To assess the potential value of R&D, organizational capability to produce R&D results should be considered and sectoral potential to exploit possible R&D outputs, including such factors as customer acceptance and availability of capital to exploit the results.

The question of alternatives is largely one of "make or buy". Another possible alternative could be to abandon R&D and the associated technological independence, become technologically dependent and buy turn-key hardware or advice. It is important to recognize that R&D strategic planning, as well as the project selection process, involves addressing program rationale and alternatives on a continuous basis. The outside experts used in an evaluation can provide a validity check on such decisions.

A major R&D evaluation question remains, related to determining the size or existence of an R&D program. On this difficult issue Blake says, in Managing for Responsive R&D:

> ...there is no precise or deterministic way of calculating the 'right' size for an R&D program... Deciding the proper size of company-funded R&D involves both quantitative and intuitive assessments ... it would be as foolish to omit quantification from a decision on R&D size as it would be impossible to omit intuition. The desirable procedure combines both intuition informed by as many quantitative facts as can be gathered and quantification adjusted by intuitive judgements of meaning and importance.

In Chapter 1, Blake gives several ideas on how to approach the problem of the size of an R&D program in the private sector. Some of these may be applicable in government operations.

This would be valuable in assessing impact but frequently cannot be used because of the nature of R&D and its contribution to programs. Of necessity, this suggested approach to evaluation focuses largely on proxy measures of impact or effects.

Application of the Approach

Before starting the evaluation, a clear understanding of its purpose should be established. The group of R&D activities to be evaluated should be specified and the information needs of decision-makers as well as the manageability of the evaluation should be considered. The focus of the evaluation of the R&D process should also be specified.

The first step should be, as always, an evaluation assessment. In this case it would focus on the R&D management process. That is, a survey for a management review should be done first. During the review, the assessor would develop an understanding of the organization, its purposes, role, environment, clients and outputs. Specifically the evaluator should collect data and analyse it as follows:

(a) Data collection:

- outline the whole management process with emphasis on strategic planning, operational planning, project selection and project review and control;
- identify R&D objectives and mandate as well as those of the supported organization;
- obtain an inventory of ongoing research activities, research personnel and facilities;
- identify the clientele for the various research activities;

- obtain a list of recent and historical research outputs and an indication of their possible effects;

- identify ongoing evaluation activities; and

- identify comparable R&D organizations.

(b) Analysis:

- This material should be analysed to help decide detailed plans for the evaluation, including reviewing management procedures and assessing research quality, possibility of impact measurement, where outside experts are needed and if so, for what aspects.

- The characteristics of the R&D being evaluated should be considered in this analysis. For example, suppose the activities consisted of a blend of targeted basic research, applied research, and long and short-term development work. Information to be developed during the evaluation could vary widely because of the different information needs of the members of the client group for the evaluation results. Top management might be interested in the appropriateness of the blend of activities, operational R&D managers in technical assessments of various projects, and the client in the work's relevance. A strategy for collecting and reporting the information developed on the evaluation should be defined at this time.

- At this point there should be a reassessment of the purpose of the evaluation, that is, can the original purpose be satisfied with the information likely to be obtained from a budget-constrained evaluation study? Can a satisfactory result be achieved with the information likely to be collected?

- A detailed plan for the execution of the evaluation should be prepared showing work to be done. The plan should take into account the characteristics of R&D, such as those listed in Appendix A.

Proposed Audit Criterion

In the OAG audit guidelines for R&D, the audit criterion for evaluation is:

> **Evaluation should be integral to the R&D management process but there should also be periodic evaluations of R&D activities conducted independently of R&D management.**

This is the standard against which we propose to conduct our audits on this issue in the future.

APPENDIX A

Characteristics of R&D That Are Relevant to its Evaluation

Long time lags. An important reason why R&D is unique and a specially designed evaluation approach is necessary arises from the long time lags that can exist between the production of a research result and the time when its impact on program achievement can be properly measured. Because of this, intermediate or operational outputs may be all that is available to serve as a proxy measure of R&D program effectiveness, and these may be unsatisfactory. Consequently a valid quantitative evaluation of the impact of R&D may be impossible.

Some of the difficulties associated with proxy measures can be seen in the example of the development of a rubella vaccine. Rubella, or measles, causes birth defects if contracted by pregnant women early in their pregnancy. Normally, people can catch rubella only once in their lives because permanent antibodies are built up during the course of the disease that prevent a second occurrence. A rubella vaccine has been developed that causes antibodies to be produced to provide some protection against subsequent contracting of the disease. A research program was involved in the development of the vaccine. The first level proxy measure would be the presence of a substance in the laboratory that had the characteristics of a vaccine. A second level proxy would be the ability of this vaccine to create in humans antibodies similar to those that the body would normally create after exposure to the rubella disease itself. However, the effectiveness of the vaccine would not become apparent until it had been ascertained that these induced antibodies were permanent and that women would not subsequently contract rubella during pregnancy. The permanence of the antibodies could only be tested after 20 or 25 years. Therefore, the effectiveness of the program might take 25 years to determine.

Uncertainty. A significant feature in R&D activities is uncertainty of outcome. In the paper we used a simple model to explain the source of some of the uncertainty. Another source is the unpredictability of progress predicated on

the creative thinking of R&D personnel. Success in creative undertakings such as research is not guaranteed, although the quality of the effort may have been high. Generally, in terms of achieving program objectives, effort counts for nothing. This is not necessarily the case with research, however, and an evaluation of R&D activity based simply on its contribution to achievement of program objectives is not always satisfactory. Risk is another side of uncertainty - where uncertainty is high, the risk of failure is high, but possible payoffs can be the greatest where the risks are highest. Some R&D activities may be expected to have a very high success rate, e.g., those involving relatively minor changes to existing processes; in others requiring extensive basic research or development, success may be infrequent but of high value. In such cases, quality of effort between infrequent successes can be a relevant measure of value: potential payoff should be considered, and value of results achieved over a short term should not be too heavily emphasized. Hence assessing the management procedure for handling uncertainty is an important topic in an evaluation of R&D.

Non-hierarchical. A manufacturing operation is intrinsically hierarchical; R&D is not. In manufacturing, from the final product there is a downward branching tree of activities and intermediate outputs needed to produce the final product. It is necessary to specify in detail for each worker particular tasks and schedule of outputs. The hierarchy of management provides the specific direction and control needed for the tree of production tasks. One of the important tools for controlling these activities is a hierarchy of increasingly more specific task objectives.

In contrast, much R&D has no natural hierarchy of tasks fitting together in a predictable way. The hierarchy of management provides only general direction, and the research workers do what they can, how they can and when they can. There are unavoidable gaps in the hierarchy of program and project objectives. For example, consider a research program designed to develop domestic products to replace imported products. Top management might state this as a priority research activity but individual research groups might not be able to allocate resources to it because they could think of no products worth working on given their particular skills and facilities.

Aggregation. R&D programs are generally collections of projects often dealing with quite separate problems in the general area being addressed. Success or failure, progress or delay in one project usually has little impact on other projects in the same program. As a result, there is usually no simple and obvious way to aggregate measures of project achievement to provide a satisfactory overall measure of R&D program achievement. Specifically, an aggregation of the effectiveness of projects is not usually an appropriate measure of the effectiveness of an R&D program. The reasons for this are many and varied but an example can show the nature of the problem.

Consider a pair of projects in an agriculture research program designed to reduce crop losses due to wheat rust. One project is to breed new strains of wheat that are resistant to continuously developing variants of wheat rust. This approach has a demonstrated history of success, but it is also expensive and relatively slow. A biochemical approach based on examining how the rust attacks the nucleus of the wheat cells could lead to the development of a specific fungicide. This approach could solve the rust problem permanently, with high potential benefits arising from reduced costs to the farmer, but the technical feasibility, environmental acceptability and marketability of the potential fungicide are questionable. An evaluation problem for research management would be to determine whether the research being conducted by the biochemist was effective in comparison with the plant breeding approach. The socio-economic benefits of the plant breeding approach are relatively obvious and easy to calculate, but the risks, costs and benefits of the research in biochemistry, which could be far from producing a candidate fungicide, are very much more difficult to assess. The problem here is that two research projects designed to attack the same problem would have to be judged differently. This situation is not unusual.

Validity of proxy measures. An evaluation of the fungicide project would have to use a distant proxy measure of effect; for example, results obtained or papers published. An evaluation of the breeding approach could use an output measure much closer to the intended effect. Further, the validity of the closer measures would be well established because the breeding approach has been used successfully before. On the other hand, the validity of papers published as a measure of progress in the fungicide project is questionable. There is a dilemma

here: the evaluator must assume the validity of such distant proxy measures although the purpose of the research is to determine whether there is a relation between current results and the desired effect. In this context, a post-evaluation is used as a pre-evaluation, and the role of outside experts is obvious.

Identifiability. The general approach to evaluating support functions like personnel or administration is to identify the program objectives and determine the contribution made to achievement of these objectives by the various support activities. R&D in most cases is a support function - its outputs are not ends in themselves. In many cases, but not all, the contribution of R&D to program objectives can be identified and the value of the contribution at least qualitatively assessed. However, in many cases, an R&D organization will be contributing to several different programs simultaneously. In a central R&D organization supporting several programs, one major function of R&D can be to provide advice on request. Such advice can be very important but cannot be quantitatively assessed. Omitting its evaluation can cause underestimation of the true value of the R&D contribution to the organization.

Heterogeneity. Typically an R&D program is made up of a collection of projects and activities differing in varying degree from each of the characteristics described above. Heterogeneity is the rule, not the exception. Consequently, most R&D programs cannot be evaluated using a standardized approach; different approaches may have to be used simultaneously.

Similarity to support functions. R&D is similar to administrative support functions in that they are not ends in themselves but means to the ends of the supported organization. A recent report published by the Economic Council of Canada* makes it clear that R&D is often a necessary but by no means sufficient factor in innovation; that is, R&D is usually best seen as a means not an end. It is generally accepted that in an evaluation of support functions it is important to determine whether they are being managed in a way that maximizes their potential contribution to the supported organization. With such functions as with R&D, management review or audit is a very important aspect of management control.

---

*K.S. Palda, B. Pazderka, Approaches to an International Comparison of Canada's R&D Expenditures. Economic Council of Canada. (1982).